Social media, financial reporting opacity and return comovement

Citation for published version:
Ding, R, Zhou, H & Li, Y 2020, 'Social media, financial reporting opacity and return comovement: Evidence from Seeking Alpha', *Journal of Financial Markets*, vol. 50, 100511.
https://doi.org/10.1016/j.finmar.2019.100511

Digital Object Identifier (DOI):
10.1016/j.finmar.2019.100511

Published In:
Journal of Financial Markets

Social media, financial reporting opacity, and return comovement: Evidence from Seeking Alpha

Rong Ding
Rong.Ding@wbs.ac.uk
Warwick Business School, University of Warwick, Coventry, CV4 7AL UK;

Hang Zhou (corresponding author)
Hang.zhou@ed.ac.uk
University of Edinburgh Business School, University of Edinburgh, Edinburgh, EH8 9JS, UK

Yifan Li
yifan.li@manchester.ac.uk
Alliance Manchester Business School, University of Manchester, Manchester, M15 6PB, UK

Abstract

In this study, we develop a model to analyze the interplay between the coverage of a firm on social media, financial reporting opacity, and stock return comovement. Our model predicts a negative association between social media coverage and comovement as social media facilitates the incorporation of firm-specific information into stock price. It also predicts that the effect of social media coverage on comovement is more pronounced among firms with higher financial reporting opacity. Using data from Seeking Alpha, the largest crowdsourced social media platform that provides "third-party generated" financial analysis in US, we find results consistent with the model's predictions.

Keywords: social media; comovement; Seeking Alpha; financial reporting opacity

JEL classification: G11, G12, G14

We are grateful to an anonymous referee, Tarun Chordia, Qiu Chen, Shujun Ding, Mingzhi Liu, Zhenyu Wu, Ivan Lim, Allaudeen Hameed and seminar participants at University of Bradford, University of Manitoba, University of Ottawa, the 2017 British Accounting and Finance Association Annual Conference, 2017 FMA Asia/Pacific Conference and the 40th European Accounting Association Annual Congress for helpful comments and suggestion.

1. Introduction

In this study, we develop a model to analyze the interplay between coverage of a firm on social media that provides financial analysis written by non-professional analysts, financial reporting opacity, and the extent to which stock return comoves with industry and market return (comovement). Both Chen et al. (2014) and Campbell et al. (2018) analyze the information conveyed by financial analysis posted on Seeking Alpha (SA), and conclude that SA articles provide reliable firm-specific information. In particular, Chen et al. (2014) show that the views expressed in SA articles can predict stock return and earnings surprises in the next three months, and such effect is more evident for articles written by contributing authors with an established track record. Campbell et al. (2018) find that the disclosure of stock positions of non-professional analysts who contribute articles to SA enhances the informativeness of their articles, which is reflected by stronger stock return surrounding their articles' publication date. From a related but different perspective, we suggest that the coverage of public firms on crowdsourced social media plays at least two roles in influencing the comovement between stock return and market and industry return.

Firstly, financial analysis written and posted on social media might have "global access" on the Internet, which facilitates the incorporation of more information, particularly firm-specific information, into stock price. Given there are an increasing number of investors who are informed after reading financial analysis posted on social media, they are able to take advantage of such information when they trade with uninformed investors. This results in firm-specific information being incorporated into stock price to a more significant extent, leading to lower return comovement.

Secondly, research shows that interaction with others in the same social network can explain a wide range of economic activities (Bikhchandani et al., 1998; Ivković and Weisbenner, 2007). In the financial markets, interaction with other market participants is critical in the dissemination of value-relevant information, as people pay more attention to ideas or facts that are reinforced by interaction, ritual, and symbols (Shiller, 1999).[1] In the Internet era, crowdsourced social media such as Seeking Alpha is becoming an important platform where individuals can exchange investment ideas with others, because it allows registered users not only to write and read articles, but also to post commentaries in response to an existing piece of information. Those who post commentaries may provide disparate perspectives or alternative insights, and can suggest corrections or even point out flaws in the original article. Not unreasonably, an article with many commentaries is expected to attract more attention from a broad audience. Therefore, the coverage on social media enables more investors to better communicate with others and understand the implications of information released in the original article, which results in such information being fully impounded into stock price, leading to reduced return comovement.

It is recognized that financial reports are an important source of firm-specific information that is widely used by investors (Bushman and Smith, 2001; Lambert, 2001). According to Jin and Myers (2006), (financial reporting) opacity, which represents the lack of information that precludes investors from determining the fair value of a firm, makes it easier for managers to conceal self-serving behavior, such as rent seeking or asset diversion.[2] This implies that the

[1] For example, Shiller and Pound (1989) find that almost all investors who recently purchased a stock had their attention drawn to it through direct communication with others.

[2] In an analytical study, Bleck and Liu (2007) posit that opacity in financial reporting constrains shareholders' ability to distinguish bad projects from good projects at an early stage, leading to the continuation of bad projects over time and a higher risk of a crash in asset prices. Such a prediction is supported by Hutton et al. (2009), who report a positive association between crash risk and the extent firms engage in accrual-based earnings management.

opacity limits the flow of firm-specific information to the market, which leads to higher return comovement. Under such circumstances, alternative information sources such as social media enable investors to access "third-party generated information" related to a firm, and use such information to assist decision-making. This suggests that social media has a stronger effect in facilitating the flow of firm-specific information to the market for firms with high financial reporting opacity, resulting in stock return of such firms being less comoving with the market and industry returns. In Section 2 we develop a model to formally derive propositions regarding the relation between social media coverage, financial reporting opacity and return comovement.

To empirically test the predictions of our model, we design a computer program to automate the process of extracting all articles published on the Seeking Alpha website between 2004 and 2014.[3] We focus on single-ticker articles that only provide information about one specific firm, and remove from our analysis all multiple-ticker articles that discuss more than one firm in one article. Then we focus on firms that have been covered by Seeking Alpha at least once in our sample period, and construct the coverage measure as the log of one plus the number of single-ticker Seeking Alpha articles for a firm during a specific year.

Our first measure of comovement is stock price synchronicity, which is defined as the extent to which variation in firm-level stock return can be explained by market and industry returns (Durnev et al., 2003; Piotroski and Roulstone, 2004; Crawford et al., 2012). Following Roll

[3] Seeking Alpha, founded in 2004 by David Jackson, is a crowdsourced social media platform that publishes financial commentary and analysis. By the end of 2015, it had 4 million registered users, while more than 10,000 registered users contribute financial commentary and analysis. Submitted articles are reviewed by a panel and are subject to editorial changes, so the quality of published articles is expected to be high. There are 7 million average monthly unique visitors and 85 million average monthly page views. Seeking Alpha has a broad coverage of stocks, including 4,000 small- and mid-cap firms. Seeking Alpha states its mission is to provide "opinion and analysis rather than news."

(1988), we measure stock price synchronicity with adjusted R^2 from the market model regression to capture the extent to which stock price movement can be explained by both market and industry-wide information.[4] After a log-transformation, a lower synchronicity measure implies that market and industry returns can explain a smaller proportion of individual stock returns, suggesting that more firm-specific information has been capitalized into stock price. Consistent with our prediction, our results confirm that Seeking Alpha coverage is associated with lower synchronicity (lower level of comovement). Such a result is economically significant, because for firms covered by Seeking Alpha, a one standard deviation increase in Seeking Alpha coverage is associated with a 4.3% reduction in the synchronicity measure.[5] Our second measure of comovement is the percentage point of a time series Pearson correlation coefficient between weekly firm return and weekly market return (*CORRE*). Based on this measure, we find consistent results that Seeking Alpha coverage is associated with lower comovement. We further corroborate that our findings are robust to estimation using a matched sample based on a propensity score matching (PSM), a two-stage least squares (2SLS) approach, and decomposing the synchronicity measure into a market return component and an industry return component. The results show that Seeking Alpha coverage is negatively associated with both the market return component and the industry return component. Taken together, our findings suggest that social media plays an important role in facilitating the flow of firm-specific information to the market, thus decreasing return comovement.

---

[4] This measure of return co-movement (and its variation) has been widely used in previous research (i.e., Morck et al., 2000; Fernandes and Ferreira, 2008; Gul et al., 2011).
[5] The reduction is calculated as percentage change in comovement before log transformation, which is the adjusted R^2 from regressing the firm level stock return on market and industry returns.

Next, we investigate whether the association between social media coverage and return comovement is more pronounced in firms with higher financial reporting opacity, because higher financial reporting opacity is expected to make it difficult for investors to obtain information from financial reports, thus making them increasingly reliant on an alternative information source, such as social media, to access firm-specific information. Consistent with Hutton et al. (2009), we use the average discretionary accruals calculated from the Francis et al. (2005) model over a three-year period as the first proxy of financial reporting opacity, and find evidence supporting the prediction of our model that the effect of social media coverage on comovement is more pronounced in firms with higher financial reporting opacity. Our inferences are qualitatively unchanged when we use analyst forecast dispersion as the second measure of financial reporting opacity, with the rationale that forecast dispersion might be larger for firms with higher degree of opacity.

The contributions of our study are threefold. Firstly, we extend the literature on the capital market consequence of social media (Blankespoor et al., 2014; Lee et al., 2015) by developing a model to explicate the relation between social media coverage, financial reporting opacity, and comovement.[6] The predictions of the model are supported by data collected from Seeking Alpha. As a typical example of crowdsourced social media specializing in financial analysis, Seeking Alpha substantially differs from other social media outlets such as Facebook or Twitter. Seeking Alpha articles are written by registered users and independent parties (i.e., the editorial team of Seeking Alpha) will verify the quality of the submission and the credentials of the author (i.e.,

[6] Blankespoor et al. (2014) find that firms can reduce information asymmetry by disseminating corporate news through Twitter, and Lee et al. (2015) show that firms use social media such as Twitter to interact with investors to attenuate the stock market's negative reaction to product recalls.

name, address, and contact information) before an article is published.[7] Furthermore, unlike tweets, which were restricted to 140 characters until September 2017, Seeking Alpha articles can accommodate the in-depth analysis of a firm, thus conveying valuable information with figures and numbers to validate the information. In particular, since January 2011, Seeking Alpha started paying each contributing author $10 per 1,000 page views.[8] It is likely, therefore, that the information in a Seeking Alpha article is of high quality, because the author has a financial incentive to publish articles with high credibility. If the information in an article is proved to be misleading, the reputation of the author will be negatively affected, and any future article from that author is unlikely to be published by Seeking Alpha. Similar incentives are absent in both Twitter and Facebook, which implies that Seeking Alpha constitutes a powerful setting to test our model. In our analysis, we find that the social media coverage has an incremental effect on return comovement in that we control for the press coverage using RavenPack data, which supports the premise that the value-relevant information provided by social media is distinctive from that released in public press.[9]

---

[7] According to an article written by Seeking Alpha CEO Ali Hoffmann and posted on its website on April 10, 2014, the editorial team of Seeking Alpha evaluates each submission based on 1) whether the idea expressed in the article is convincing; 2) whether the idea is actionable; and 3) whether the idea is well-presented (https://seekingalpha.com/article/2134803-how-much-does-seeking-alpha-pay-its-contributors).

[8] According to the same article written by Seeking Alpha CEO Ali Hoffmann and posted on its website on April 10, 2014, Seeking Alpha pays authors who contribute articles exclusive to Seeking Alpha. The base payment is $10 for each 1,000-page views. For high-quality analysis of stocks that lack good research (e.g., small-cap), Seeking Alpha pays a minimum $150 for articles selected by its editors, and $500 for top small-cap ideas with exceptionally attractive risk/reward profiles.

[9] RavenPack collects and analyses business news from news providers such as Dow Jones newswire, the *Wall Street Journal*, *Barron's*, industry and business publications, regional newspapers, and regulatory updates. Research shows that public information from RavenPack is distinctive from information on social media. For example, Giannini et al. (2019) measure investors' divergence of opinion with the difference between sentiment of news articles from the Dow Jones Factiva news database and sentiment of stock Twitter messages from Stocktwits.com, and find that the divergence of sentiments is associated with a greater trading volume on and after an earnings announcement. Cookson and Niessner (2019) use the sentiment of messages collected from Stocktwits.com to construct dispersion of investors' opinion, and find the dispersion measure is distinct from other factors that influence trading volume including investors' attention and news coverage about the firm. These findings are consistent with views from public news being different from those based on social media communication.

Secondly, our study provides new insights into the determinants of stock return comovement. Since Roll (1988), a considerable amount of research has identified links between comovement and investor protection and the development of financial markets, corporate governance, mandatory adoption of XBRL, and newspaper coverage (Morck et al., 2000; Durnev et al., 2003; Crawford et al., 2012; Dong and Ni, 2014; Dong et al., 2016). Our study points to the role of social media in influencing stock return comovement.

Finally, our results highlight that social media coverage has an incremental effect on the incorporation of firm-specific information into stock price for firms with higher financial reporting opacity, which suggests that social media can complement formal disclosure (e.g., financial statements) to a certain extent in unravelling value-relevant information to the market. Our findings have implications for the executives of public firms, because firms could coordinate with social media platforms to facilitate the dissemination of corporate information and provide an independent verification and evaluation of such information to a broader audience.

Our study is related to but different from Filzen and Schutle (2017), who use Google search volume as a proxy of investors' information demand, and show that an increase in financial reporting complexity (measured by the length of 10-Q reports) is followed by an increase in Google search and subsequently a significant increase in comovement between the focal firm and its peers. In contrast, our study concentrates on the information supply captured by social media coverage and the impact of improved information accessibility to investors on the comovement between the firm and the market. In particular, our results that the influence of social media coverage on comovement is more pronounced for firms with higher financial reporting opacity echo the findings in Filzen and Schutle (2017) because firms with greater market frictions (in terms of either reporting opacity or reporting complexity) benefit more from improved corporate information environment. Finally, Filzen and Schutle (2017) measure return comovement using the R^2 of the regression of a firm's stock return on the average return of its four closest peers (firms with a similar predicted enterprise value to sales ratio), while we measure comovement with as adjusted R^2 of a regression of a firm's return on both market and industry returns. Therefore, our findings corroborate the role social media plays in enhancing market efficiency and resource allocation in a relatively broad sense.[10]

Our study is also different from Dasgupta et al. (2010). Firstly, in Dasgupta et al. (2010), their view of financial reporting transparency is associated with events of disclosing general firm-specific information. Therefore, they study major events through which firms actively disclose a big chunk of information (e.g., seasoned equity offerings or cross-listing events). However, these are publicly available firm-specific events that are accessible to all investors, and are likely to be captured by the RavenPack dataset. In contrast, we measure financial reporting opacity with the quality of information from financial statements that can be understood by investors with sufficient skill, which reflects the quality of more technical firm-specific information. Secondly, Dasgupta et al. (2010) are interested in a dynamic relation between transparency and comovement, while we focus on a contemporaneous and static relation between the two. The different research design between our study and Dasgupta et al. (2010) enables us to explore how

[10] We thank an anonymous reviewer for drawing our attention to Filzen and Schutle (2017).

the firm-level variation in financial reporting quality changes the importance of social media coverage in influencing the stock return comovement.

The remainder of the paper is organized as follows. We develop the model in Section 2. Sample and research design are described in Section 3, and Section 4 presents the empirical results. Concluding remarks are in the final section.

2. The model

In the literature, security prices are considered to be jointly determined by a common market factor and an idiosyncratic factor (Roll, 1988; Jin and Myers, 2006).[11] When a larger fraction of price fluctuation is caused by the idiosyncratic factor, the comovement between asset return and the market factor decreases. In contrast, if the market factor drives the price fluctuation to a greater extent, it is reasonable that asset returns will cointegrate with the market factor, generating higher return comovement. Our two-period model, which is adapted from the models proposed by Schmidt (2012), Kacperczyk et al. (2016), and Huang et al. (2018), provides insights into how social media coverage and financial reporting opacity relate to investors' learning about shocks of asset payoffs, which consequently influence the return comovement. In this section, we briefly outline the key intuitions and conclusions of our model for conciseness and refer the readers to Appendix 1 for a complete exposition of the model.

In our two-period model, we assume there exists $N+1$ assets in the market: N individual stocks and a composite stock representing the market index. The payoffs of all assets are determined by

[11] Roll (1988) suggests that the idiosyncratic factor is largely determined by firm-specific information.

a market-wide shock, while individual assets are also affected by firm-specific payoff shocks. We assume that there exists a continuum of atomless investors in the market. Investors have the same prior information about the payoffs of assets at the start of the first period but can learn and refine their beliefs about these shocks, which will be revealed at the end of the second period. During the first period, investors receive signals about the payoff of the asset and rebalance their portfolio optimally based on the signals they receive.

At the end of the first period, investors maximize their final utility by choosing the optimal demand of each security based on their posterior beliefs about each asset conditioning on the signals, which further determines the prices of each asset. Return comovement between an individual stock and the market can therefore be examined as the correlation between the first-period returns of an individual asset and the composite asset.

Our contributions to the models in Schmidt (2012), Kacperczyk et al. (2016), and Huang et al. (2018) are twofold. Firstly, we consider the effect of social media coverage and financial reporting opacity on investors' learning process about asset payoffs. In our model, media coverage and financial reporting opacity are directly associated with the precisions of the firm-specific signals that investors can receive. Intuitively, investors can obtain more (less) precise signals about firm-specific shocks with higher (lower) social media coverage and lower (higher) financial reporting opacity. Secondly, we account for the fact that firm specific information disclosed on social media is relatively costless and accessible compared to that contained in financial reports, which requires specific skills to interpret. In our model, we assume that investors are heterogeneous: only a proportion of the investors are “skilled” such that they are able to interpret the information disclosed in the financial reports and obtain a more precise firm-specific signal compared to the "unskilled" investors.

Based on our model, we derive two important propositions:

Proposition 1: Comovement between firm-specific return and market return is smaller for firms with a higher level of social media coverage.

Proposition 2: The marginal effect of social media coverage on return comovement is more pronounced for firms with higher financial reporting opacity.

In the remainder of this paper, we empirically test Propositions 1 and 2 derived from our theoretical two-period model using fixed-effect regression models. Details of the design of the tests are in Subsection 3.2.

3. Data, variable descriptions, and research design

3.1 Data

Our measure of social media coverage is the number of articles related to a firm for a given year (single-ticker articles) that are posted on Seeking Alpha, the largest crowdsourced social media platform in the United States. We design a computer program to extract all single-ticker articles from the Seeking Alpha website. Specifically, we use a python program based on "Scrapy" to extract all single-ticker articles from the website in HTML format. Our dataset includes 133,217 single-ticker articles between 2004 and 2014. In our analysis, we use the natural log of one plus the number of Seeking Alpha articles to alleviate its skewness. Appendix 2 provides an example of a typical Seeking Alpha article. We collect return data from CRSP, firm fundamental data from Compustat, and analyst coverage data from I/B/E/S. We require non-missing values[12] for key variables including the synchronicity and control variables listed in Subsection 3.2. Our final sample contains 39,568 firm-year observations from between 2004 and 2014.[13] To mitigate the potential influence of outliers, we winsorize all continuous variables at the top and bottom 1%.

3.2 Research design

3.2.1 Test of Seeking Alpha coverage and stock return comovement

Roll (1988) is the first to propose that stock price synchronicity, the association between a firm's stock return and market and industry returns, is negatively associated with the amount of firm-specific information being impounded into stock prices. Following Huang et al. (2018), we use two measures for the stock price synchronicity. The first measure is the adjusted R^2 from the following regression for each firm-year:

$$RET_{i,t} = \alpha_0 + \beta_1 \mathrm{Re}t\_mkt_{i,t} + \beta_2 \mathrm{Re}t\_ind_{i,t} + \beta_3 \mathrm{Re}t\_mkt_{i,t-1} + \beta_4 \mathrm{Re}t\_ind_{i,t-1} + \varepsilon_{i,t}\,. \quad (1)$$

where *RET* is the weekly stock return of individual firm *i* in week *t*; *Ret_mkt* is the weekly return calculated as the value-weighted return on all NYSE, AMEX, and NASDAQ stocks in week *t*; *Ret_ind* is the weekly return of the industry (based on the two-digit SIC code) in week *t* to which the firm belongs. The lag returns are included to account for non-synchronous trading. Adjusted R^2 is derived from equation (1). We run Regression (1) across each firm-year with a minimum of 45 weekly observations. Following the literature (i.e., Morck et al., 2000; Piotroski and Roulstone, 2004), we define synchronicity as:

[12] We keep the entire sample of U.S. stocks irrespective of the existence of Seeking Alpha coverage to avoid the endogenous nature of Seeking Alpha coverage.

[13] The number of observations used to test Propositions 1 and 2 varies between 39,568 and 12,606, as the data requirements we impose to calculate CORRE and measures of financial reporting opacity result in the loss of some of the observations. However, when we investigate Propositions 1 and 2 with a reduced sample including the same number of observations, our inferences remain qualitatively unchanged.

$$SYNCH = \log(\frac{Adj.R^2}{1 - Adj.R^2}) \text{ .} \tag{2a}$$

The benefit of log transformation of $Adj.R^2$ is the creation of an unbounded variable out of a variable originally bounded between 0 and 1, which generates a dependent variable with approximately normal distribution. By construction, higher value of stock price synchronicity (*SYNCH*) indicates that firms' stock returns are closely tied to market and industry returns (higher return comovement).

Our second measure of comovement is the percentage point of the time series Pearson correlation coefficient between weekly firm return and weekly market return, which is also used in Peng and Xiong (2006) and Anton and Polk (2014). For the time series return of firm *i*, R_i, and market return R_m, *CORRE* is computed as follows:

$$CORRE_{i,m} = \frac{COV(R_i, R_m)}{\sigma_{Ri}\sigma_{Rm}}. \tag{2b}$$

To test the association between Seeking Alpha coverage and return comovement, we estimate the following model:

$$\begin{aligned} Comovement_{i,t} &= \alpha_0 + \beta_1 L\_SA_{i,t} + \beta_2 L\_RP_{i,t} + \beta_3 LNUM_{i,t} + \beta_4 SIZE_{i,t} + \beta_5 LMB_{i,t} + \beta_6 LEVERAGE_{i,t} \\ &+ \beta_7 ROA_{i,t} + \beta_8 NIND_{i,t} + \beta_9 HERFSALE_{i,t} + \beta_{10} STDROA_{i,t} + \beta_{11} BIG4_{i,t} + \sum \alpha_i Firm_i + \sum \alpha_j Year_j + \varepsilon_{i,t} \end{aligned} \text{ .} \tag{3}$$

where *Comovement* is either stock price synchronicity or *CORRE* for firm *i* in year *t*, and $L\_SA_{i,t}$ is defined as the natural log of one plus the number of Seeking Alpha single-ticker articles covering firm *i* in year *t*.

To distinguish the effect of Seeking Alpha coverage on the stock price comovement from that of public firm-specific news, we use press coverage data provided by RavenPack Analytics (RP) as a control variable. RavenPack collects and processes articles from premium newswires, providers of regulatory news and press releases in real time for both firm-specific and macroeconomic-related news globally. For each firm-specific article, it identifies the entity involved in the article and the relevance of the entity to the article. From RavenPack, we collect all firm-specific news articles for U.S. firms that are considered as the most relevant within our sample period.[14] We then compute $L\_RP_{i,t}$ as the natural log of one plus the number of the most relevant articles covering firm *i* in year *t*.

We also incorporate a set of control variables that have been identified as affecting stock price synchronicity (Roll, 1988; Hutton et al., 2009; Crawford et al., 2012; Kim and Shi, 2012). *LNUM* is constructed as the natural log of 1 plus the average number of analysts following the firm during the previous fiscal year[15]. *SIZE* is defined as the natural log of the firm's market capitalization at the end of the previous fiscal year; *LMB* is measured as the natural log of market capitalization scaled by the book value of equity at the end of the previous fiscal year; *LEVERAGE* is the total long-term debt scaled by the total assets at the end of the previous fiscal year; and *ROA* is measured as income before extraordinary items divided by total assets at the end of the previous fiscal year. *NIND* is the natural log of the number of firms in the industry to which firm *i* belongs; *HERFSALE* is the sum of the squared terms of the proportion of a firm's

[14] RavenPack assigns a relevance score from 0 to 100 to each article to evaluate the relevance of the article to the entities involved. The article will receive a relevance score of 100 if a) it is found to be playing a key role in the first event detected in the headline of a story, and b) it is not playing an explicitly lower relevant role such as a rater. We only keep articles with a relevance score of 100 to enhance the power of our test.

[15] We also incorporate the residual analyst coverage into the model, which is obtained by regressing analyst coverage on firm size. The results are qualitatively the same. We thank an anonymous reviewer for raising this issue.

revenue to total revenue in the industry at the end of the previous fiscal year; and *STDROA* is the standard deviation of *ROA* in the previous five years. To capture the impact of the quality of external auditing, *Big4* is set to 1 if the firm is audited by one of the Big 4 accounting firms (PwC, Deloitte, Ernst & Young or KPMG), 0 otherwise. We adjust the standard error for heteroscedasticity, as well as serial and cross-sectional correlation using a two-dimensional cluster at the firm and year level (Petersen, 2009). Finally, we include firm fixed effects and year fixed effects to address firm-specific and time series trends of stock price synchronicity. As suggested by Dyreng et al. (2010), the utilization of firm fixed effects forces the firm to act as its own control, and our test essentially concentrates on within-firm variation. The definition of all variables is in Appendix 3.

3.2.2 The moderating effect of financial reporting opacity

To measure financial reporting opacity, we first follow Hutton et al. (2009) to calculate the three-year average accruals quality over year *t*-2 to year *t*, and label it *OPA*. Discretionary accruals are computed as the residual from the estimation error model (equation (4)). Following Francis et al.'s (2005) model, we calculate the absolute value of the residual from each year for the two-digit SIC industry, and larger discretionary accruals (larger absolute value of residual) indicate lower accrual quality:

$$\frac{TACC_{i,t}}{TA_{i,t-1}} = \lambda_0 + \frac{\lambda_1 CFO_{i,t-1}}{TA_{i,t-1}} + \frac{\lambda_2 CFO_{i,t}}{TA_{i,t-1}} + \frac{\lambda_3 CFO_{i,t+1}}{TA_{i,t-1}} + \frac{\lambda_4 \Delta \mathrm{Re}\, v_{i,t}}{TA_{i,t-1}} + \frac{\lambda_3 PPE_{i,t}}{TA_{i,t-1}} + \varepsilon_{i,t} \,. \tag{4}$$

Our second measure of financial reporting opacity is the dispersion of analyst earnings forecasts. Research (Maffett, 2012) shows that accounting information asymmetry is associated with greater analyst forecast dispersion, as the disagreement of earnings forecasts among analysts reflects the level of information asymmetry between corporate insiders and outsiders (i.e., investors and intermediaries such as analysts). Opaque firms disclose less firm-specific information or information of inferior quality, which makes it difficult for analysts to reach a consensus on earnings forecasts, generating large forecast dispersion. We compute the dispersion of analyst earnings forecasts for each firm in year *t*, and create *OPA2* by scaling the dispersion with the firm's opening stock price of the year.

Then we employ the model in equation (5) to test whether Seeking Alpha coverage plays a more pronounced role in decreasing return comovement among firms with higher financial reporting opacity:

$$\begin{aligned} Comovement_{i,t} = {} & \alpha_0 + \beta_1 L\_SA_{i,t} + \beta_2 OPA_{i,t} + \beta_3 L\_SA_{i,t} * OPA_{i,t} + \beta_4 L\_RP_{i,t} + \beta_5 LNUM_{i,t} + \beta_6 SIZE_{i,t} + \beta_7 LMB_{i,t} \\ & + \beta_8 LEVERAGE_{i,t} + \beta_9 ROA_{i,t} + \beta_{10} NIND_{i,t} + \beta_{11} HERFSALE_{i,t} + \beta_{12} STDROA_{i,t} + \beta_{13} BIG4_{i,t} + \sum \alpha_i Firm_i + \sum \alpha_j Year_j + \varepsilon_{i,t} \end{aligned} \quad . \qquad (5)$$

where *Comovement* is either stock price synchronicity or *CORRE* for firm *i* in year *t*. We run equation (5) for the full sample, and expect the coefficient of interaction between Seeking Alpha coverage and *OPA* (*OPA2*), our main variable of interest, to be significantly negative.

4. Empirical results

4.1 Descriptive statistics

Figure 1 presents the summary statistics on the Seeking Alpha coverage. The number of Seeking Alpha articles (firms covered by Seeking Alpha) increased from 18 (12) in 2004 to 21,995 (2,931) in 2014, showing the substantially growing influence of Seeking Alpha in the investment community during our sample period.[16]

---

[16] The statistics are calculated after excluding financial assets apart from stocks such as mutual funds, ETFs, and non-U.S. stocks from the original full sample of Seeking Alpha single-ticker articles.

Table 1 provides the descriptive statistics of all the variables. The synchronicity measure has a mean (median) of -2.664 (-2.421), and varies from -3.874 (25th percentile) to -1.234 (75th percentile). *CORRE* has a mean (median) of 40.898 (43.003) and varies from 24.361 (25th percentile) to 58.662 (75th percentile). The mean (median) of the number of Seeking Alpha articles (*L_SA*) is 0.426 (0.000). The mean (median) of the number of news articles from RavenPack (*L_RP*) is 4.612 (5.517), indicating that there are on average a significantly larger amount of publicly available news articles than Seeking Alpha articles. The mean (median) of analyst following is 0.952 (0.000), and mean (median) of firm size measured by the logarithm of market capitalization is 6.213 (6.184), which suggests our sample is populated with large firms. *ROA* has a mean (median) of 0.025 (0.024), and standard deviation of *ROA* has a mean (median) of 0.101 (0.033). Finally, the mean (median) of *Big4* is 0.701 (1.000), indicating that more than 70% of our sample firms have a Big 4 firm as their auditor. All the variables have substantial variation.

<< Insert Table 1 about here >>

4.2 Correlation

Table 2 presents the Pearson correlations between the variables. Both *CORRE* and the synchronicity measure are positively correlated with Seeking Alpha coverage, which seems inconsistent with our prediction.[17] However, as we do not control for the other determinants of the synchronicity measure, the correlation has to be interpreted with caution. Consistent with the findings of Piotroski and Roulstone (2004), the two comovement measures are positively correlated with analyst following, suggesting that high analyst coverage facilitates the disclosure

[17] The positive correlation likely captures the relation between synchronicity measure and firm size (correlation = 0.645, $p < 0.001$), which in turn is positively associated with Seeking Alpha coverage.

of market and industry information to the market. Seeking Alpha coverage is weakly and positively correlated with the public press coverage (from RavenPack), which signals that Seeking Alpha authors are more likely to cover firms with higher news coverage. The two comovement measures are positively correlated with the number of RP news articles, *SIZE*, market-to-book, *LEVERAGE*, *ROA, BIG4*, and is negatively related to the standard deviation of *ROA* and *NIND*. Finally, the correlation statistics do not raise concerns regarding multicollinearity, as the largest correlation is that between *CORRE* and *SYNCH* (0.847) and they will not enter the same regression. The VIF of all subsequent regressions are below 10.

<< Insert Table 2 about here >>

4.3 Multivariate results

Table 3 reports results related to the prediction that Seeking Alpha coverage is associated with lower return comovement. We use the logarithm of one plus the number of articles on Seeking Alpha (*L_SA*) as the proxy of Seeking Alpha coverage, and synchronicity (columns (1) and (2)) and *CORRE* (column (3) and (4)) are used as proxy of comovement. In the regression for column 1(3), we regress synchronicity (*CORRE*) on Seeking Alpha coverage, as well as analyst coverage and firm size as control variables, because the correlations between these two variables and the synchronicity measure are the highest among the correlations between all control variables and the synchronicity measure. In the regression for column 2(4), we employ the complete set of control variables in the analysis, and also control for both firm-fixed and time-fixed effects. As the results across the columns are consistent, we focus on the interpretation of the results in columns (2) and (4). When synchronicity and *CORRE* are the dependent variables, the coefficient of *L_SA* is negative and significant (-0.057*,* *t* = -3.946*;* -0.462*;* *t* = -2.415 respectively), which indicates that more Seeking Alpha coverage enables the incorporation of firm-specific information into stock price to a greater extent, leading to lower comovement. The results lend credence to the contention that investment-related information disclosed on social media can effectively be revealed to the market and capitalized into stock price. Furthermore, we calculate the marginal effect to gauge the economic significance. With all other variables at their sample mean, a standard deviation increase of *L_SA* from its mean results in a 4.3% (9.50%) reduction in stock price synchronicity (*CORRE*). Therefore, we conclude that the negative association between Seeking Alpha coverage and return comovement is both statistically and economically significant. We further decompose the synchronicity measure into a market return component and an industry return component, and in untabulated analysis find that Seeking Alpha coverage is negatively associated with both components with statistical significance. It is important to note that our results are obtained after controlling for the public new release from RavenPack, which indicates that Seeking Alpha coverage contains firm-specific information that is different from what is released in the public news.

For the rest of the control variables, the coefficients of analyst coverage and size in Table 3 are significantly positive, whereas the coefficients of market-to-book are negative and significant. These findings are broadly consistent with Piotroski and Roulstone (2004) and Crawford et al. (2012).

<< Insert Table 3 about here >>

Table 4 shows results related to the prediction that the influence of Seeking Alpha coverage on return comovement is more pronounced in firms with high financial reporting opacity. We use the three-year average of discretionary accruals calculated from the Francis et al.'s (2005) model (*OPA*), and analyst forecast dispersion (*OPA2*) as proxies for financial reporting opacity, so firms with large average discretionary accruals (larger forecast dispersion) are considered to be more opaque. We introduce an interaction term between *OPA* (*OPA2*) and the Seeking Alpha coverage measures (*L_SA*), and the coefficient of the interaction term is the main variable of interest.

We present four models, where *OPA* (*OPA2*) is the proxy of opacity in Models 1 and 2 (3 and 4). In columns (1) and (2) in Table 4, where synchronicity and *CORRE* are the dependent variables in the regression, the coefficient of the interaction between *L_SA* and *OPA* is negative and significant (-0.142, t = -4.297; -1.938, t = -3.650 respectively), In columns (3) and (4) where synchronicity and *CORRE* are the dependent variables in the regression, the coefficient of interaction between *L_SA* and *OPA*2 is significantly negative (-0.073, t = -4.711; -0.451, t = -2.695 respectively). The findings suggest that relative to firms with lower opacity (smaller discretionary accruals and smaller forecast dispersion), Seeking Alpha coverage plays a more significant role in reducing return comovement for firms with high opacity (larger discretionary accruals and larger forecast dispersion). It is likely that investors (in particular individual investors) find it difficult and costly to acquire information for firms with high opacity, and Seeking Alpha articles effectively enable the flow of firm-specific information on such firms to the market, leading to an incremental decrease in return comovement.[18] Our results are

[18] We divide our sample into low opacity and high opacity subsamples based on a sample mean of discretionary accruals and analyst forecast dispersion, and run the baseline model in the two subsamples. Untabulated results

consistent with the prediction of our model that the marginal effect of opacity is less than that of Seeking Alpha coverage. This is likely due to the small number of skilled investors who are capable of extracting information from financial reports, which diminishes the marginal effect of opacity on return co\movements.[19]

<< Insert Table 4 about here >>

4.4 Endogeneity issue

4.4.1 Propensity score matching

Our results may suffer from endogeneity, because firms are less likely to be randomly covered by social media such as Seeking Alpha. For example, large firms, firms with extreme unexpected earnings or firms in select industries (i.e., consumer-oriented industries) are more likely to be covered on Seeking Alpha. Our first approach to address the concern of endogeneity is the propensity score matching (PSM) method. We estimate the following logit model for each year: the dependent variable is coded 1 if a firm is covered by Seeking Alpha in a given year and zero otherwise; the independent variables include all firm-level control variables in equation (3). Secondly, with replacement we match each "treatment firm" (a firm covered by Seeking Alpha in a given year *t*) with two matching firms (firms not covered by Seeking Alpha in the same year) that have the closet propensity scores within a maximum distance of 1%. That is, we use a nearest-neighbor matching approach with common support and a caliper constraint of 0.01. We have 21,528 observations for this analysis. The matching appears successful as the standardized biases of variables are less than 5% after the matching. We include year fixed effect in the first

---

show that both the magnitude and significance levels of the Seeking Alpha coverage measure are higher in the high-opacity subsample, suggesting that Seeking Alpha coverage plays a more significant role in reducing return comovement among firms with high opacity. This is consistent with the findings based on regression analysis incorporating interaction between the measure of opacity and Seeking Alpha coverage.

[19] The measurement of investor skill is beyond the scope of our study, so we leave further investigation of this issue for future research.

stage prediction model, which effectively removes the time trend of increasing Seeking Alpha coverage over the sample period.

We repeat the analysis using the PSM sample, and the results are reported in Table 5. It is clear that the tenor of our results remains qualitatively unchanged, because the coefficient of Seeking Alpha coverage is significantly negative (-0.053, t = -3.013; -1.085, t = -4.899) in regressions when synchronicity and *CORRE* are used as the comovement measure.

<< Insert Table 5 about here >>

4.4.2 Two-stage least squares

We employ two instrumental variables (IVs) for Seeking Alpha coverage to further mitigate the issue of endogeneity. The first instrument is the annual advertising expenditures of a firm. We construct the instrumental variable *L_ADX* using the natural log of 1 plus the firm's annual advertising expenditure. The second instrument is the intensity of Seeking Alpha coverage for the industry to which a firm belongs. The second instrumental variable, *L_SA_IND*, is measured as the natural log of 1 plus the total annual number of Seeking Alpha articles for the industry to which the firm belongs. We expect Seeking Alpha coverage to be positively correlated with both IVs, because firms with higher advertising expenditures are more likely to attract the attention of both investors in general and registered users of Seeking Alpha in particular. Firms belonging to industries that attract more attention from Seeking Alpha are more likely to be covered. On the other hand, firm-level advertising expenditures and industry-level Seeking Alpha coverage are less likely to have a direct influence on return comovement at the firm level. Both IVs pass the over-identification test, and the results are consistent with our prediction. In the first stage, we find that both IVs are significantly and positively associated with Seeking Alpha coverage at the firm level. In the second stage, the negative and significant effect of predicted Seeking Alpha coverage on comovement remains, thus corroborating our findings reported in Subsection 4.3.

<< Insert Table 6 about here >>

4.4.3 "Day of the Week" effect

We test the "day of the week effect" based on the conjecture that articles published on weekdays (Monday-Friday) attract more attention from investors, who are able to trade in response to the information released in these articles in a timely manner. In contrast, investors could be less attentive to articles published on weekends, causing a delayed or insignificant reaction to information released in weekend articles.[20] Therefore, our prediction is that articles published on Saturday and Sunday have a relatively smaller effect on comovement. We test such a conjecture with Model 6, with the expectation that β_1 (the coefficient of weekend coverage) is significantly smaller than β_2 (the coefficient of weekday coverage).

$$\begin{aligned} Comovement_{i,t} &= \alpha_0 + \beta_1 L\_SA\_WEEKENDS_{i,t} + \beta_2 L\_SA\_WEEKDAY_{i,t} + \beta_3 L\_RP_{i,t} + \beta_4 LNUM_{i,t} \\ &+ \beta_5 SIZE_{i,t} + \beta_6 LMB_{i,t} + \beta_7 LEVERAGE_{i,t} + \beta_8 ROA_{i,t} + \beta_9 NIND_{i,t} + \beta_{10} HERFSALE_{i,t} + \beta_{11} STDROA_{i,t} \\ &+ \beta_{12} BIG4_{i,t} + \sum \alpha_i Firm_i + \sum \alpha_j Quarter_j + \varepsilon_{i,t} \end{aligned} \quad (6)$$

L_SA_WEEKENDS is the natural log of 1 plus the number of single-ticker Seeking Alpha articles posted on Saturday and Sunday for a firm in the quarter. *L_SA_WEEKDAY* is the natural log of 1

[20] Such prediction is consistent with findings in prior research. For example, DellaVigna and Pollet (2009) compare the reaction to earnings announcement on Friday with the reaction to announcement on other days in the week, and find that immediate stock response is 15% lower while the delayed response is 70% larger for Friday announcement. They interpret the findings as consistent with investors being more distracted on Friday because of the incoming weekend, which leads to delayed incorporation of new information into stock price.

plus the number of single-ticker Seeking Alpha articles posted from Monday to Friday for a firm in the quarter. The results in Table 7 are consistent with our prediction (β_1 = -0.106, t = -3.682; β_2 = -0.177, t = -11.223 when synchronicity is the dependent variable; β_1 = -1.652, t = -5.192; β_2 = -2.159, t = -13.469 when *CORRE* is the dependent variable). F-tests confirm that in both regressions, β_1 is significantly smaller than β_2.[21]

<< Insert Table 7 about here >>

4.5 Robustness check

In this subsection, we first test whether Seeking Alpha coverage facilitates the incorporation of a type of firm-specific information, future earnings, into the current stock price. We use the following model, which is outlined in Kothari and Sloan (1992):

$$RET_{i,t,t-k} = \rho_0 + \rho_1 * E_{i,t} + \varepsilon_{i,t} . \quad (7)$$

where $RET_{i,t,t-k}$ is the stock return from period *t-k* to period *t*. $E_{i,t}$ is the earnings in period *t* (future earnings), defined as income before extraordinary items divided by total assets. As predicted by Kothari and Sloan (1992), when the time interval *k* increases, firm-level information becomes more likely to be incorporated into the return over the *t–k* to *t period*. Hence, ρ_1 should increase with *k*. In the case where investors retrieve firm-level information earlier, the estimated ρ_1 will be larger when the estimated interval is longer. We estimate the following system of equations using the seemingly unrelated regression (SUR) technique, as SUR provides more efficient estimation than separate OLS regressions when the disturbances of the equations are related.

$$RET_{i,t,t-1} = \rho_0 + \rho_1 EARNING_{i,t} + \rho_2 EARNING_{i,t} * L\_SA_{i,t-1} + Controls + \varepsilon_{i,t} . \quad 8(a)$$

21 Our results remain consistent when we classify Monday-Thursday (Friday-Sunday) as weekday (weekend).

$$RET_{i,t,t-2} = \theta_0 + \theta_1 EARNING_{i,t} + \theta_2 EARNING_{i,t} * L\_SA_{i,t-2} + Controls + \varepsilon_{i,t}. \quad 8(b)$$

The ratio of θ_2 in equation 8(b) to ρ_2 in equation 8(a) measures the relative speed with which stock price incorporates future earnings for firms with higher Seeking Alpha coverage. The larger the ratio, the earlier investors incorporate a firm's future earnings into the current stock price for firms with higher Seeking Alpha coverage. Evidence supporting such prediction is consistent with the notion that higher Seeking Alpha coverage is associated with more forward-looking information being capitalized into the current stock price.[22]

Table 8 presents results supporting the view that Seeking Alpha coverage facilitates the incorporation of future earnings into current stock price. In particular, θ_2 in the two-year period estimation (0.051*,* t = 6.284) is significantly larger than ρ_2 in the one-year estimation (0.006*,* t = 1.626). The difference of coefficients results in a Wald statistic of 23.64 (p < 0.01). To summarize, we find evidence that the stock prices of firms with higher Seeking Alpha coverage incorporate future earnings more efficiently.

<< Insert Table 8 about here >>

As mentioned earlier, we include all available firm-year observations irrespective of Seeking Alpha coverage, as an attempt to alleviate potential endogenous selection issues of Seeking Alpha articles. However, the cost of our chosen sample selection criteria is the possibility that

[22] We acknowledge that when the measurement window is lengthened to capture the return-earnings relation, firm-specific information that was previously concealed could be eventually disclosed through alternative sources such as analyst reports or management disclosure. But this would work against us finding a significant effect of social media coverage to facilitate the incorporation of earnings-related information into stock price. We thank an anonymous reviewer for raising this issue.

our test results may be driven by large firms with high investor attention.[23] Therefore, we re-examine the baseline model results across different size groups of our sample. That is, we partition our sample into high and low size sub-samples based on the annual median values of *SIZE*, and compare the baseline regression coefficients of *L_SA*.

We maintain the same set of control variables as in the early analysis. The results are presented in Table 9. The coefficients for *L_SA* are of statistical significance among both large firms (-0.046, t = -3.145; -0.500, t = -2.402) and small firms (-0.052, t = -1.765; -1.034, t = -3.053), suggesting that our baseline results are not primarily attributable to large-sized firms. Our results further confirm that the Seeking Alpha platform serves as a supplement to the conventional public media coverage, as RavenPack coverage plays an insignificant role among smaller stocks (0.032, t = 0.854; *0.787*, t = 1.813).

<< Insert Table 9 about here >>

We conduct additional robustness checks. For brevity, we do not tabulate these results. As Seeking Alpha articles that receive more commentaries would attract more attention, we classify articles into influential ones and less influential ones based on the median value of commentaries on the original article for a given year, and expect influential articles to play a more significant role in reducing return comovement. The untabulated results confirm our prediction that influential Seeking Alpha articles have more effect on reducing return comovement than less influential articles.

5. Conclusion

[23] This issue has also been reflected by the high correlation between *SIZE* and Seeking Alpha coverage.

Mindful of the increasing importance of social media as a venue for information production and dissemination in the new millennium, in this paper we develop a model to predict that the coverage of a public firm on social media enables investors to access credible and precise information about a firm. Consequently, this facilitates the transmission of more firm-specific information into stock price, resulting in lower return comovement. We test the prediction of our model with data collected from the Seeking Alpha website for the 2004-2014 period, and find that Seeking Alpha coverage is negatively associated with return comovement. In addition, we show that the effect of Seeking Alpha coverage on return comovement is more salient in firms with higher financial reporting opacity. Our findings are robust to propensity score matching, a two-stage least squares (2SLS) approach, and alternative measures of firm opacity.

Our study is subject to an important caveat. As suggested by Chen et al. (2014) and Campbell et al. (2018), Seeking Alpha, a popular venue for both professional and non-professional investors to share their research output, represents a credible source of information. Seeking Alpha is different from other social media platforms that allow users to post information without any verification of the information (e.g., Twitter). As our analysis is built on data collected from Seeking Alpha, we recommend that readers cautiously generalize the inferences of the current study to settings based on other social media platforms.

Our study is of interest to investors because sophisticated financial market participants might be incentivized to develop trading techniques that consider the coverage on social media when formulating their trading strategy. Our findings have important implication for executives and regulators, as the social media landscape shifting in that it has become a revolutionary approach to information generation, evaluation, and dissemination due to its global access and interactive nature.

References

Anton, M. and Polk, C. (2014), Connected Stocks. *Journal of Finance*, 69: 1099-1127.

Barberis, N., A. Shleifer, and J. Wurgler. (2005). Co-movement. *Journal of Financial Economics*, 75(2), 283–317.

Beuselinck, C., P. Joos, I. Khurana, and S. Van der Meulen. (2010). Mandatory IFRS reporting and stock price informativeness. *Working paper*, Tilburg University.

Bikhchandani, S., D. Hirshleifer, and I. Welch. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *Journal of Economic Perspectives*, *12*, 151–170.

Blankespoor, E., G. Miller, and H. White. (2014). The role of dissemination in market liquidity: Evidence from firms' use of Twitter. *Accounting Review*, *89*, 79–112.

Bleck, A., and X. Liu. (2007). Market transparency and the accounting regime. *Journal of Accounting Research*, *45*, 229–256.

Bushman, M., and J. Smith. (2001). Financial accounting information and corporate governance. *Journal of Accounting and Economics*, *32*, 237–333.

Campbell, J., DeAngelis, M., and Moon, J. (2018). Skin in the game: personal stock holdings and investors' response to sock analysis on social media. *Working paper*, University of Georgia.

Chen, H, P. De, Y. Hu, and B. Hwang. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*, 1367–1403.

Cohen, L., and D. Lou. (2012). Complicated firms. *Journal of Financial Economics*, *104*, 383–400.

Collins, D. W., S. P. Kothari, J. Shanken, and R. Sloan. (1994). Lack of timeliness and noise as explanation for the low contemporaneous return–earnings association. *Journal of Accounting and Economics*, *18*, 289–324.

Cookson, J., and M. Niessner (2019). Why don't we agree? Evidence from a social network of investors. *Journal of Finance*, Forthcoming.

Crawford, S., D. Roulstone., and E. So. (2012). Analyst initiations of coverage and stock price efficiency. *Accounting Review*, *87*, 1527–1553.

Dasgupta, S., J, Gan., and N, Gao. (2010). Transparency, price informative-ness and stock return synchronicity: Theory and evidence. *Journal of Financial and Quantitative Analysis, 45*, 1189-1220.

Dechow, M., G. Sloan, and P. Sweeney. (1995). Detecting earnings management. *Accounting Review*, *70*, 193–225.

DellaVigna, S., and Pollet, J. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance, 64*, 709-749.

Dong, Y., and C. Ni. (2014). Does limited attention constrain investors' acquisition of firm-specific information? *Journal of Business Finance and Accounting*, *41*, 1361–1392.

Dong, Y., O. Z. Li, Y. Lin, and C. Ni. (2016). Does information processing cost affect firm-specific information acquisition? Evidence from XBRL adoption. *Journal of Financial and Quantitative Analysis*, *51*, 435–462.

Drake, M., J. Thornock, and B. Twedt. (2017). The Internet as an information intermediary. *Review of Accounting Studies, 22,*1-34.

Durnev, A., R. Morck, B. Yeung, and P. Zarowin. (2003). Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research*, *41*, 791–836.

Dyreng, S., M. Hanlon, and E. Maydew, 2010. The effects of executives on corporate tax avoidance. *Accounting Review*, *85*, 1163–1189.

Fernandes, N. and M. Ferreira. (2008). Does international cross-listing improve the information environment? *Journal of Financial Economics*, *88*, 216–244.

Fernandes, N., and M. Ferreira. (2009). Insider trading laws and stock price informativeness. *Review of Financial Studies*, *22*, 1845–1887.

Ferreira, M., and P. Laux. (2007). Corporate governance, idiosyncratic risk, and information flow. *Journal of Finance*, *62*, 951–989.

Filzen, J., and M, Schutte. (2017). Comovement, financial reporting complexity and information markets: evidence from the effect of changes in 10-Q length on internet search volumes and peer correlations. *North American Journal of Economics and Finance*, 39, 19-37.

Francis, J., R. LaFond, P. Olsson, and K, Schipper. (2005). The market pricing of accruals quality. Journal of accounting and economics, 39(2), 295-327.

Giannini, R., Irvine, P., and T, Shu. (2019). The convergence and divergence of investors' opinions around earnings news: evidence from a social network. *Journal of Financial Markets*, 42, 94-120.

Grossman, S., and J. Stiglitz. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*, 393–408.

Gul, F. A., B. Srinidhi, and A. C. Ng. (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics*, *51*, 314–338.

Haggard, S., X. Martin, and R. Pereira. (2008). Does voluntary disclosure improve stock price informativeness? *Financial Management*, *37*, 747–768.

Haw, I., B. Hu, J. J. Lee, and W. Wu. (2012). Investor protection and price informativeness about future earnings: International evidence. *Review of Accounting Studies*, *17*, 389–419.

Hirshleifer, D., Teoh, S. (2003). Limited attention, information disclosure and financial reporting. *Journal of Accounting and Economics*, 35, 337-386.

Hirshleifer, D., Lim, S., Teoh, S. (2003). Disclosure to a credulous audience: the role of limited attention. *Working paper*, The Ohio State University.

Hope, O. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research*, *41*, 235–272.

Huang, S., Y. Huang, and T.-C. Lin (2018). Attention allocation and return co-Movement: Evidence from repeated natural experiments. *Journal of Financial Economics*, Forthcoming.

Hutton, P., A. Marcus, and H. Tehranian. (2009). Opaque financial reports, R^2, and crash risk. *Journal of Financial Economics*, *94*(1), 67–86.

Ivković, Z., and S, Weisbenner. (2007). Information diffusion effects in individual investors' common stock purchases: Covet the neighbors' investment choices. *Review of Financial Studies*, *20*, 1327–1357.

Jin, L., and S. Myers. (2006) R^2 around the world: New theory and new tests. *Journal of Financial Economics*, *79*, 257–292.

Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L. (2016) A Rational Theory of Mutual Funds' Attention Allocation, Econometrica, *84(2),* 571-626

Kim, J., and H. Shi. (2012). IFRS reporting, firm-specific information flows, and institutional environments: International evidence. *Review of Accounting Studies*, *17*, 474–517.

Kothari, S. P., and R. Sloan. (1992). Information in prices about future earnings. *Journal of Accounting and Economics*, *15*, 143–171.

Lambert, A. (2001). Contracting theory and accounting. *Journal of Accounting and Economics*, *32*, 3–87.

Lee, F., A. Hutton, and S. Shu. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, *53*, 367–404.

Li, K., R. Morck, F. Yang, and B. Yeung. (2004). Firm-specific variation and openness in emerging markets. *Review of Economics and Statistics*, *86*, 658–669.

Maffett, M. (2012). Financial reporting opacity and informed trading by international institutional investors. *Journal of Accounting and Economics*, *54*, 201–220.

Morck, R., B. Yeung, and W. Yu. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial* Economics, 58, 215–260.

Peng, L and Xiong, W (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, *80(3),* 563-602.

Petersen, M. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22, 435–480.

Piotroski, J., and D. Roulstone. (2004). The influence of analysts, institutional investors and insiders on the incorporation of market, industry and firm-specific information into stock prices. *Accounting Review*, 79, 1119–1151.

Roll, R. (1988). R^2. *Journal of Finance*, 43, 541–566.

Schmidt, D. (2012) Investors' Attention and Stock Covariation: Evidence from Google Sport Searches. Working Paper, INSEAD

Shiller, R. (1995). Conversation, information and herd behavior. American Economic Review, 85, 181–185.

Shiller, R. (1999). Human behavior and the efficiency of the financial system. In J. B. Taylor and M. Woodford (eds.), Handbook of Macroeconomics, 1, 1305–1340.

Shiller, R., and J. Pound. (1989). Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior and Organization*, 12, 47–66.

Tucker, J. W., and P. Zarowin. (2006). Does income smoothing improve earnings informativeness? *Accounting Review*, 81, 251–270.

Yu, J. (2011). Stock price informativeness and corporate governance: An international study. *International Review of Finance*, 11, 477–514.

Appendix 1: The two-period model

Our two-period model is developed based on the theoretical framework of Schmidt (2012), Kacperczyk et al. (2016), and Huang et al. (2018). We provide insights on how social media coverage and financial reporting opacity relate to the comovement between firm-specific returns and market (industry) returns.

1.1 General setting

Assets. In the model we consider an economy of two periods: $t = 0,1,2$. At time 0, investors have no prior beliefs about the payoffs of the securities. At time 1, investors receive signals about the payoff of each stock. They then update their beliefs and rebalance their portfolio correspondingly. The payoffs of the assets are realized at time 2.

We assume there are $N+1$ assets in the market. The first N stocks are firm-specific stocks, and the $(N+1)$ -th stock is a composite stock denoted using the subscript c , which is understood as a stand-in for all other assets in the space of assets considered (the whole market or industry)[24]. The payoffs of the assets at time 2 are specified as follows:

$$\begin{aligned} &f_i = \mu + c + e_i, \quad f_c = \mu + c, \quad i \in \{1,\ldots,N\}, \\ &e_i \sim \mathrm{N}\,(0,\sigma_i), \quad c \sim \mathrm{N}\,(0,\sigma_c), \end{aligned} \tag{A.1}$$

where μ is the mean payoff of all stocks. The random variable c is a common payoff shock to all assets, and e_i is the firm-specific payoff shock to f_i. The composite stock allows us to properly define the return synchronicity as the squared correlation between the returns of an asset to the composite asset. For simplicity, we assume that $\sigma_c = \sigma_i = \sigma > 0$ and that c , e_i , and e_j are independent for all $i \neq j$.

[24] Note that one can also add another set of assets that controls for sector-wise shocks within the given space of assets. However, our conclusions on the comovement between firm-specific return and the composite return of the whole asset space remain valid if we replace the composite return with a sector-wise return. We therefore omit the results for sector-wise return for conciseness.

At time $t=0$, all investors have a prior belief about the distribution of the payoff vector $f=\{f_1,\ldots,f_N,f_m\}'$. Based on our assumption in (A.1), we see that $E_0[f]=\mu\iota$ and $V_0[f]=\Sigma$, where we denote $\iota=\mathbf{1}_{N+1\times 1}$, and $E_t[x]$ and $V_t[x]$ are understood as the mean and variance function of the random variable x conditional on time t. The variance-covariance matrix Σ of the payoff vector is:

$$\Sigma=\iota\iota'\sigma+diag\{\sigma,\ldots,\sigma,0\}. \tag{A.2}$$

In addition to the $N+1$ risky assets, there exists a risk-free asset with the payoff normalized as 1 for each unit.

Investors. We assume a continuum of investors. The j-th investor is indexed by an element in the unit interval $j\in[0,1]$. Investors at time $t=0$ will trade these assets based on their prior knowledge. At time $t=1$, investors receive noisy signals about each shock, and will adjust their beliefs about each asset accordingly. Deviating from Schmidt (2012) and Kacperczyk et al. (2016), we assume that the precision of the signals for each stock is given, as we are not interested in the attention allocation of the investors. Instead, we assume that the maximum precision of the signal for a stock is determined by the information available to the investors, as we elaborate below.

Signals about c and e_i that the j-th investor receive are specified as follows:

$$s_{c,j}\sim\mathrm{N}\left(c,\frac{1}{E_{c,j}}\right),\quad s_{i,j}\sim\mathrm{N}\left(e_i,\frac{1}{E_{i,j}}\right), \tag{A.3}$$

In this specification, $E_{c,j}$ and $E_{i,j}$ are the precision of the composite and firm-specific signals that the j-th investor receives. Intuitively, the larger these variables are, the more precise the associated signals become. If any precision variable is equal to zero, then the signal has infinite variance and does not help improve the prior beliefs about the associated asset.

By setting $E_{i,j}$ to be different across j, we also allow heterogeneity among investors, which we interpret as the skill level of investors. However, different from Kacperczyk et al. (2016), who define the skill level of investors based on whether they can refine a noisy signal, we differentiate the two types of investors based on the type of firm-specific information they can exploit to refine e_i. Specifically, we assume that for each individual stock there exist two types of information: free and processed information, such as social media coverage, and unprocessed or technically demanding information such as the financial reports for a firm.

We consider Seeking alpha coverage and the transparency of financial reports as two representatives for processed and unprocessed information, which is denoted by two positive constants, α_i and O_i respectively. Higher α_i and O_i indicate more social media coverage and more transparent financial reports, which translates into more unique processed and unprocessed information about the firm-specific signal that further generates a more precise signal, and vice versa. We denote the skill level of the j-th investor as π_j, which equals one if she is skilled, and zero otherwise. We denote the proportion of skilled investors by $\chi\%$. The signal precision of the i-th stock for the j-th investor is therefore:

$$E_{i,j} = \alpha_i + O_i \pi_j. \tag{A.4}$$

Intuitively, for an individual stock, signal precision for a skilled investor will be higher than that of an unskilled investor whenever $O_i > 0$ because a skilled investor can exploit extra unprocessed information in the financial reports that an unskilled investor is not able to decipher precisely. For the composite shock, we assume that $E_{c,j} = E_c$ for all investors, as the precision of the composite shock is likely to be improved by publicly available news such as macroeconomic announcements that are available to all investors.

Following Schmidt (2012) and Huang et al. (2018), we assume that investors do not learn from stock prices, and the supply of stocks is non-random and normalized to be ι for all t. However, we note that adding a stochastic component to the supply of stocks will not alter our results qualitatively.

Bayesian Learning. At $t=1$, each investor observes signal realizations $s_{c,j}$ and $s_{i,j}$. We can write the signal about each asset compactly in a matrix notation:

$$s_j \equiv \iota\mu + \{s_{1,j}+s_{c,j},\ldots,s_{N,j}+s_{c,j},s_{c,j}\}' \sim \mathrm{N}\ (f,\Sigma_{s,j}), \tag{A.5}$$

where:

$$\Sigma_{s,j} = \iota\iota'\frac{1}{E_c} + diag\{\frac{1}{E_{1,j}},\ldots,\frac{1}{E_{N,j}},0\}. \tag{A.6}$$

The j-th investor learns from the signal s_j to adjust their belief about the assets using Bayes' law. Since both f and s_j are multivariate normal, it follows that the conditional distribution $f\,|\,s_j$ is also normally distributed:

$$f\,|\,s_j \sim \mathrm{N}\ ((\Sigma^{-1}+\Sigma_{s,j}^{-1})^{-1}(\Sigma^{-1}\iota\mu+\Sigma_{s,j}^{-1}s_j),(\Sigma^{-1}+\Sigma_{s,j}^{-1})^{-1}). \tag{A.7}$$

We denote $\hat{\Sigma}_j = (\Sigma^{-1}+\Sigma_{s,j}^{-1})^{-1}$ as the posterior variance of the assets. By denoting $\hat{\sigma}_{i,j}^{-1} = \sigma^{-1}+E_{i,j}$ and $\hat{\sigma}_c^{-1} = \sigma^{-1}+E_c$, we can write $\hat{\Sigma}_j$ in the following compact form:

$$\hat{\Sigma}_j = \iota\iota'\hat{\sigma}_c + diag\{\hat{\sigma}_{1,j},\ldots,\hat{\sigma}_{N,j},0\}. \tag{A.8}$$

Consequently, we can also write the posterior mean vector for the j-th investor as:

$$\hat{\mu}_j = \hat{\Sigma}_j(\Sigma^{-1}\iota\mu+\Sigma_{s,j}^{-1}s_j). \tag{A.9}$$

Portfolio choice and equilibrium price. Each investor has an initial wealth of W_0, and we assume that each investor has a mean-variance utility function with a mean reversion parameter $\rho > 0$: $U_{t,j} = \rho E_t[W_{2,j}] - \frac{\rho^2}{2} V_t[W_{2,j}]$. We denote the total wealth the j-th investor has at time t as $W_{t,j}$. At time $t = 0$, each investor will choose the amount they want to trade for each stock, $q_{0,j}$, by solving the following optimization problem:

$$\begin{aligned} &\max_{q_{0,j}} \ \rho E_0[W_{2,j}] - \frac{\rho^2}{2} V_0[W_{2,j}], \\ &s.t. \quad W_{2,j} = W_0 + q'_{0,j}(p_1 - p_0) + q'_{1,j}(f - p_1), \end{aligned} \tag{A.10}$$

where p_t is the price vector for all assets at time t, and $q_{1,j}$ is the amount of stocks traded at time $t = 1$.

At time $t = 1$, each investor receives signals about each asset and will rebalance their portfolio by solving the following optimization problem:

$$\begin{aligned} &\max_{q_{1,j}} \ \rho E_1[W_{2,j}] - \frac{\rho^2}{2} V_1[W_{2,j}], \\ &s.t. \quad W_{2,j} = W_{1,j} + q'_{1,j}(f - p_1). \end{aligned} \tag{A.11}$$

Given the conditional distribution of f in (A.7), we can solve the maximization problem inside the expectation of (A.11). This is in fact a well-studied problem in classic finance literature. For the j-th investor, the optimal quantity to trade $q_{1,j}$ is:

$$\hat{q}_{1,j} = \frac{1}{\rho} \hat{\Sigma}_j^{-1} (\hat{\mu}_j - p_1). \tag{A.12}$$

Now the market clearing condition implies that:

$$\int \hat{q}_{1,j} dj = \iota. \tag{A.13}$$

Integrating over j on both sides of (A.12) yields:

$$\rho\iota = \int \hat{\Sigma}_j^{-1}\hat{\mu}_j dj - p_1 \int \hat{\Sigma}_j^{-1} dj, \tag{A.14}$$

where p_1 is understood as the equilibrium vector of asset prices at time $t=1$. Substituting (A.9) into equation (A.14), we have:

$$\rho\iota = \Sigma^{-1}(\iota\mu - s_j) + \int \hat{\Sigma}_j^{-1} s_j dj - p_1 \int \hat{\Sigma}_j^{-1} dj. \tag{A.15}$$

The aggregated inverse posterior variance-covariance matrix is just a weighted average of the inverse posterior variance-covariance matrix from two types of investors:

$$\overline{\Sigma}^{-1} \equiv \int \hat{\Sigma}_j^{-1} dj = \chi\overline{\Sigma}_\pi^{-1} + (1-\chi)\overline{\Sigma}_{1-\pi}^{-1}, \tag{A.16}$$

where $\overline{\Sigma}_\pi^{-1} \equiv \int \hat{\Sigma}_j^{-1}\pi_j dj$ is the aggregated inverse posterior variance-covariance for all skilled traders, and $\overline{\Sigma}_{1-\pi}^{-1}$ is defined analogously for the unskilled traders. Substituting (A.16) into (A.15) and using that[25] $\int \hat{\Sigma}_j s_i dj = \overline{\Sigma} f$ and $\int s_j dj = f$, we can write the equilibrium price vector p_1 in a compact form:

$$p_1 = \overline{\mu} - \rho\overline{\Sigma}\iota, \tag{A.17}$$

where $\overline{\mu} \equiv (I - \overline{\Sigma}\Sigma^{-1})f + \mu\overline{\Sigma}\Sigma^{-1}\iota$ is understood as the aggregate posterior mean vector and I is the identity matrix.

1.2 Return comovement

In this paper, we focus on the contemporaneous marginal effects of α_i and O_i on the comovement between firm-specific return and composite return. Therefore, following the approach in Schmidt (2012) and Huang et al. (2018), we focus on the return from $t=0$ to $t=1$, formally defined as $r = p_1 - p_0 = \{r_1, \ldots, r_N, r_c\}'$. From Appendix A.1 in Schmidt (2012), p_0 can be expressed as follows:

[25] These relations are true because $E_0[s_j] = f$ and that signal errors are uncorrelated with signal precisions.

$$p_0 = \mu - \rho \Sigma \iota. \quad \text{(A.18)}$$

As a result, the stock return vector from $t = 0$ to $t = 1$ is defined as:

$$r = p_1 - p_0 = (I - \bar{\Sigma}\Sigma^{-1}) f - \mu + \rho(\bar{\Sigma} - \Sigma)\iota. \quad \text{(A.19)}$$

Therefore, the variance of r is of the following form:

$$\begin{aligned} V_0[r] &= (\Sigma - \bar{\Sigma})\Sigma^{-1}(\Sigma - \bar{\Sigma}) \\ &= \iota\iota'(\bar{\sigma}_c - \sigma)^2 \sigma^{-1} + \sigma^{-1} diag\{(\bar{\sigma}_1 - \sigma)^2, \ldots, (\bar{\sigma}_N - \sigma)^2, 0\}, \end{aligned} \quad \text{(A.20)}$$

where $\bar{\sigma}_c$ and $\bar{\sigma}_i$ are the aggregated posterior variance for the composite and individual assets, correspondingly. Using that all investors have the same precision for the composite shock and that investors of the same skill level have the same precision for each firm-specific shock, we find that:

$$\begin{aligned} \bar{\sigma}_c^{-1} &= \sigma^{-1} + E_c, \\ \bar{\sigma}_i^{-1} &= \int \hat{\sigma}_{i,j}^{-1} dj = \sigma^{-1} + \alpha_i + \chi O_i. \end{aligned} \quad \text{(A.21)}$$

Consequently, we can define the return comovement as the squared Pearson correlation coefficient between r_i and r_c:

$$R_i^2 = \frac{Cov_0^2(r_i, r_c)}{V_0[r_i]V_0[r_c]} = \frac{1}{1 + \left(\frac{\sigma - \bar{\sigma}_i}{\sigma - \bar{\sigma}_c}\right)^2}. \quad \text{(A.22)}$$

For mathematical convenience, we work with the following monotone transformation of R_i^2, which corresponds to the synchronicity measure defined in equation (2a):

$$SYNCH_i = \ln\left(\frac{R_i^2}{1 - R_i^2}\right) = 2\ln\left(\frac{\sigma - \bar{\sigma}_c}{\sigma - \bar{\sigma}_i}\right). \quad \text{(A.23)}$$

Substituting (A.21) into (A.24), we see that:

$$SYNCH_i = 2\ln\left(1 - \frac{1}{1 + \sigma E_c}\right) - 2\ln\left(1 - \frac{1}{1 + \sigma(\alpha_i + \chi O_i)}\right). \quad \text{(A.24)}$$

Based on (A.24), we deduce Propositions 3 and 4 below, which support Propositions 1 and 2 in Section 2 from a theoretical perspective.

Proposition 3. For two individual stocks *m* and *n* with the same level of O_m and O_n, if $\alpha_m > \alpha_n$, then $SYNCH_m < SYNCH_n$.

Proof. Taking the partial derivative of $SYNCH_i$ w.r.t. α_i, we have:

$$\frac{\partial SYNCH_i}{\partial \alpha_i} = -\frac{2}{(\alpha_i + \chi O_i)(\sigma(\alpha_i + \chi O_i) + 1)} < 0. \qquad \text{(A.25)}$$

Therefore, it is immediate that $SYNCH_i$ is a monotonically decreasing function of α_i holding O_i constant. ■

Interestingly, the marginal effect of O_i on $SYNCH_i$ is much weaker. By taking the partial derivative of $SYNCH_i$ w.r.t. α_i and O_i, we see that:

$$\frac{\partial SYNCH_i}{\partial O_i} = -\frac{2\chi}{(\alpha_i + \chi O_i)(\sigma(\alpha_i + \chi O_i) + 1)} < 0. \qquad \text{(A.26)}$$

This shows that the marginal effect of O_i on $SYNCH_i$ is only $\chi\%$ of that of α_i. If the percentage of skilled traders is small, then in empirical estimation the marginal effect of O_i can be swamped by the estimation noise and appear insignificant. Intuitively, the marginal effect of O_i on $SYNCH_i$ is only present for the skilled investors, since the unskilled investors do not react to an increase in O_i. As a result, a contemporaneous shift in the transparency of financial reports will always have a much weaker effect on the stock return comovement than Seeking Alpha coverage unless all investors are able to correctly interpret financial reports and incorporate this information into stock return.

Proposition 4. For two individual stocks *m* and *n* with the same level of α_m and α_n, $\left|\frac{\partial SYNCH_m}{\partial \alpha_m}\right| > \left|\frac{\partial SYNCH_n}{\partial \alpha_n}\right|$ if $O_m < O_n$.

Proof. Proposition 4 is equivalent to the claim that $\frac{\partial^2 SYNCH_i}{\partial \alpha_i \partial O_i} > 0$ since the first order derivatives are negative according to (A.25). Straightforward derivation shows that:

$$\frac{\partial^2 SYNCH_i}{\partial \alpha_i \partial O_i} = \frac{2\chi(2\sigma(\alpha_i + \chi O_i) + 1)}{(\alpha_i + \chi O_i)^2 (\sigma(\alpha_i + \chi O_i) + 1)^2} > 0. \tag{A.27}$$

This completes the proof. ■

It is evident that Propositions 3 and 4 lead to Propositions 1 and 2 directly. Moreover, we can test for these propositions empirically using fixed effect panel regressions, as marginal effects are captured by regression coefficients by construction.

Appendix 2: Seeking Alpha article example

Amazon Earnings Broadly As Expected

Apr. 25, 2013 5:39 PM ET
135 comments About: Amazon.com, Inc. (AMZN)
Paulo Santos
(10,045 followers)
Long/short equity, arbitrage, event-driven

Amazon (NASDAQ:AMZN) reported its Q1 2013 earnings. These came in at $0.18 versus a $0.09 consensus. At first the stock climbed quite a bit on the notion that it had beat or doubled expectations, but one needs to consider that for Amazon, $0.10 in excess or missing on its earnings is basically irrelevant, because it needs just $46 million or a puny 0.28% of sales for a beat or miss of that magnitude.

Also predictable, Amazon's revenues came in slightly below consensus ($16.07 billion vs $16.16 billion consensus). More relevant was Amazon's guidance for Q2 2013, which was as follows:

Net sales are expected to be between $14.5 billion and $16.2 billion, or to grow between 13% and 26% compared with second quarter 2012.

Operating income (loss) is expected to be between $(340) million and $10 million, compared to $107 million in the comparable prior year period.

This guidance includes approximately $340 million for stock-based compensation and amortization of intangible assets, and it assumes, among other things, that no additional business acquisitions, investments, or legal settlements are concluded and that there are no further revisions to stock-based compensation estimates.

As I predicted before the earnings were released, this constitutes another guide-down for Amazon's revenues. The midpoint of the guidance falls at $15.35 billion whereas present consensus sits at $15.94. I'd expect consensus to be revised lower to around $15.7-$15.8 billion or so.

Comparison to my model 1

The model 1 predictions compared as follows to what Amazon actually reported:

($ million)	Model 1 Q1 2013	Actual Q1 2013
Revenues		
Product	13161	13271
Services	2782	2799
total revenues	**15943**	**16070**
COGS product	11714	11801
Gross margin	4229	4269
Gross margin % of revenues	**26.5%**	**26.6%**
Fulfillment	1716	1796
Marketing	624	632
Technology	1464	1383
G&A	265	246
Other	25	32
total operating costs	**15807**	**15890**
Operating income	**136**	**180**
% of revenues	**0.85%**	**1.12%**
Interest income	10	10
Interest expense	-28	-33
Other	-20	-77
total	**-38**	**-100**
Income before taxes	**98**	**80**
Income taxes	-34	18
Equity-method	-25	-17
Net income	**38**	**82**
Diluted shares (million)	463.0	463.0
EPS	**$0.08**	**$0.18**

* Using gross margin from GMV

In what regards my own modelling, where I use my model 1 for both short term and long term predictions, the major differences were in 3 cost lines and 1 margin line:

- **Product margins** came in at 11.3% versus my 11.0% assumption. My 2013 assumption is 11.1%, which I will revise towards 11.2%;

- **Technology**, which came in 5.5% below my estimate. This implied a ratio of Technology/Other revenue of 173.3% versus my Q1 2013 assumption of 183% … but it should be noted that my 2013 assumption is 175% so lower than Q1 2013. I will revise my long term assumptions down 2% per year as a result;
- **G&A**, which came in 7.2% below my estimate. This implied a ratio of G&A/GMV of 0.78% versus my assumption of 0.85%. Since my 2013 assumption is already 0.80% this will mean no change as this number is somewhat volatile and the yearly assumption is already below Q1 and near the realised value;
- **Fulfillment**, which came in 4.7% above my estimate. This implied a ratio of Fulfillment/GMV of 5.71% versus my assumption of 5.50%. Q1 is usually the lowest in this regard so this implies a higher 2013 assumption. Presently the assumption is at 5.64%, so I will change the model towards 5.7%.

My own long-term model already implies that technology will get better (less costly) over the long-term, so no surprise there. G&A has some volatility so it won't imply much of a change. As for fulfilment, it might have negative implications for the long term.

All in all the cost relationships held quite well. The minor $40 million difference in net profit is well within the kind of uncertainty one can expect while predicting a company of Amazon's size and basically came from the product margins being slightly ahead of expectations, probably still from the higher margins enjoyed by the new Kindle Fires.

It should also be noted that every revenue growth assumption was very close to what Amazon reported, from 1P to 3P to other revenue.

Revised long-term model

Taking into account the slight differences explained, my revised long-term model now predicts the following:

Amazon.com P&L ($ million)

	2012	2013	2014	2015	2016	2017	2018	2019	2020
Revenues		20.4%	18.7%	16.7%	14.4%	13.1%	11.6%	10.0%	8.8%
Product	51733	60528	69607	78656	87308	96039	104682	113057	120971
Services	9360	13004	17687	23224	29291	35868	42592	48980	55348
total revenues	61093	73531	87293	101880	116598	131907	147274	162037	176318
COGS product	45971	53749	61811	69846	77529	85282	92958	100394	107422
Gross margin using GMV	15124	19375	24068	29353	35084	41213	47500	53636	59693
Gross margin	15122	19783	25483	32034	39069	46625	54316	61643	68896
Gross margin % of revenues	24.8%	26.9%	29.2%	31.4%	33.5%	35.3%	36.9%	38.0%	39.1%
Fulfillment	6419	8181	10162	12394	14813	17401	20056	22646	25204
Marketing	2408	3014	3744	4566	5458	6411	7389	8343	9286
Technology	4564	6763	9558	12560	15794	19076	22091	24484	26627
G&A	896	1148	1426	1739	2079	2442	2815	3178	3537
Other	159	100	100	100	100	100	100	100	100
total operating costs	60417	72954	86800	101205	115773	130712	145408	159146	172175
Operating income	676	577	493	675	825	1195	1866	2891	4143
% of revenues	1.11%	0.79%	0.56%	0.66%	0.71%	0.91%	1.27%	1.78%	2.35%
Interest income	40								
Interest expense	-92								
Other	-80	-150							
total	-132	-200	-200	-200	-200	-200	-200	-200	-200
Income before taxes	544	377	293	475	625	995	1666	2691	3943
Income taxes	-428	-82	-102	-166	-219	-348	-583	-942	-1380
Equity-method	-155	-100	-100	-100	-100	-100	-100	-100	-100
Net income	-39	195	90	209	306	547	983	1649	2463
Diluted shares (million)	470.0	473.8	478.4	483.6	489.7	496.5	504.2	512.6	521.8
EPS	-$0.08	$0.41	$0.19	$0.43	$0.63	$1.10	$1.95	$3.22	$4.72

-++

The predictions are unchanged for the most part, with the margins and cost lines basically compensating each other, only the lower tax rate ends up having a slight positive effect for 2013.

Conclusion

Amazon's earnings report brought nothing new. The growth rates and costs continue mostly as expected - the cost relationships held, with most uncertainty remaining on technology, where improvement is already expected and always difficult to model.

These cost relationships mean that Amazon will have a lot of trouble ever meeting the lofty expectations the Street has for it. At the same time Amazon's growth rates continue to falter and perhaps somewhat amazingly, net shipping costs increased again.

There was nothing in the report to change my opinion that Amazon is a clear short which will never produce enough profit to justify the levels it trades at. I expect this report to lead to another round of downward estimate revisions in terms of revenues, and perhaps also in terms of EPS. These revisions are systematic because the long-term models the Street uses do not respect the stable cost relationships that I have identified,

Disclosure: I am short AMZN. I wrote this article myself, and it expresses my own opinions. I am not receiving compensation for it (other than from Seeking Alpha). I have no business relationship with any company whose stock is mentioned in this article.

Appendix 3: Variable definitions

Variable	Definition
SYNCH	Stock return synchronicity after log transformation.
CORRE	Percentage points of time-series Pearson correlation coefficient between weekly firm return and weekly market returns.
L_SA	Natural log of (SA article number + 1).
L_RP	Natural log of (number of most relevant articles from RavenPack + 1).
LNUM	Natural log of (number of analyst coverage + 1).
SIZE	Natural log of (Firm's market capitalization).
LMB	Natural log of (market capitalization scaled by the book value of equity).
LEVERAGE	Total long term debt scaled by total assets.
ROA	Income before extraordinary items scaled by total assets.
NIND	Natural log of the number of firms in the industry to which firm I belongs.
HERFSALE	Sum of squared terms of the proportion of a firm's revenue to total revenue in the industry.
STDROA	Standard deviation of the ratio between income before extraordinary items and total asset in the previous five years.
BIG4	Dummy variable, set to 1 if the firm is audited by one of the Big 4 audit firms.
OPA	Measure of financial opacity, which is the average of accrual quality value of the previous three years. Accrual quality measure is based on the Francis et al. (2005) model. Measured as absolute value of residual from each year two-digit SIC industry. cross sectional regression: $\frac{TAcc_{i,t}}{TA_{i,t-1}} = \alpha_0 + \frac{\beta_1 CFO_{i,t-1}}{TA_{i,t-1}} + \frac{\beta_2 CFO_{i,t}}{TA_{i,t-1}} + \frac{\beta_3 CFO_{i,t+1}}{TA_{i,t-1}} + \frac{\beta_4\ \Delta Rev_{i,t}}{TA_{i,t-1}} + \frac{\beta_5\ PPE_{i,t}}{TA_{i,t-1}} + \varepsilon_{i,t}$
OPA2	Analyst earnings forecast dispersion scaled by firm's opening stock price of the year.
EARNING	Earnings before extraordinary items scaled by total assets.
L_SA_IND	Natural log of (the total annual SA coverage of the industry to which the firm belongs + 1).
L_ADX	Natural log of (firm's annual advertising expenditure + 1).

Figure 1: Summary statistics of Seeking Alpha articles

Year	**Seeking Alpha article**	**Firms covered by Seeking Alpha**
2004	18	12
2005	797	283
2006	4271	1203
2007	10264	1925
2008	9337	1843
2009	9957	1759
2010	9528	1934
2011	10794	2008
2012	12273	2146
2013	14782	2370
2014	21995	2931

Table 1: Descriptive statistics

Variable	Mean	Std. Dev.	Min	P25	Median	P75	Max	*N*
SYNCH	-2.664	1.962	-16.689	-3.874	-2.421	-1.234	5.508	39568
CORRE	40.898	23.194	-45.684	24.361	43.003	58.662	94.502	39568
L_SA	0.426	0.788	0.000	0.000	0.000	0.693	6.632	39568
L_RP	4.613	2.451	0.000	4.466	5.517	6.205	7.843	39568
LNUM	0.952	1.108	0.000	0.000	0.000	1.923	4.022	39568
SIZE	6.213	2.119	-1.038	4.661	6.184	7.670	13.348	39568
LMB	0.702	0.883	-3.332	0.176	0.634	1.150	9.241	39568
LEVERAGE	0.154	0.173	0.000	0.001	0.097	0.255	0.969	39568
ROA	0.025	2.308	-10.655	-0.008	0.024	0.067	226.310	39568
NIND	5.157	1.195	0.000	4.263	5.389	6.194	6.796	39568
HERFSALE	0.075	0.080	0.010	0.035	0.045	0.090	1.000	39568
STDROA	0.101	0.960	0.000	0.013	0.033	0.087	92.564	39568
BIG4	0.701	0.458	0.000	0.000	1.000	1.000	1.000	39568
OPA	0.071	0.156	0.000	0.024	0.042	0.080	8.641	22847

Table 1 presents the descriptive statistics of the variables. The sample contains 39,568 firm-year observations over the period 2004–2014. P25 (P75) is the 25th (75th) percentile of the variable's distribution. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm returns and the weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm *i* belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise; *OPA* is the previous five years' standard deviation of accrual quality value based on the Francis et al. (2005) model.

Table 2: Correlation table

	SYNCH	*CORRE*	*L_SA*	*L_RP*	*LNUM*	*SIZE*	*LMB*	*LEVERAGE*	*ROA*	*NIND*	*HERFSALE*	*STDROA*
CORRE	0.847***											
L_SA	0.262***	0.222***										
L_RP	0.153***	0.174***	0.137***									
LNUM	0.410***	0.393***	0.398***	0.240***								
SIZE	0.630***	0.592***	0.464***	0.190***	0.541***							
LMB	0.085***	0.086***	0.143***	0.082***	0.154***	0.336***						
LEV	0.195***	0.171***	0.058***	0.062***	0.099***	0.235***	0.083***					
ROA	0.018***	0.019***	0.015***	-0.022***	0.006	0.037***	0.115***	-0.010				
NIND	-0.153***	-0.109***	-0.100***	-0.059***	-0.134***	-0.127***	0.028***	-0.181***	0.011*			
HERFSALE	0.090***	0.061***	0.065***	0.028***	0.095***	0.053***	-0.028***	0.082***	-0.006	-0.740***		
STDROA	-0.031***	-0.027***	0.003	-0.027***	-0.035***	-0.035***	0.131***	-0.036***	0.814***	0.030***	-0.017***	
BIG4	0.408***	0.402***	0.203***	0.100***	0.315***	0.544***	0.161***	0.197***	0.019***	-0.167***	0.086***	-0.005

Table 2 reports the Pearson correlation of variables used in the analysis. The sample contains 39,568 firm-year observations over the period 2004–2014. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm return and weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm *i* belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise; *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Baseline model

$$Comovement_{i,t} = \alpha_0 + \beta_1 L\_SA_{i,t} + \beta_2 L\_RP_{i,t} + \beta_3 LNUM_{i,t} + \beta_4 SIZE_{i,t} + \beta_5 LMB_{i,t} + \beta_6 LEVERAGE_{i,t}$$
$$+\beta_7 ROA_{i,t} + \beta_8 NIND_{i,t} + \beta_9 HERFSALE_{i,t} + \beta_{10} STDROA_{i,t} + \beta_{11} BIG4_{i,t} + \sum \alpha_i Firm_i + \sum \alpha_j Year_j + \varepsilon_{i,t}.$$

VARIABLES	*SYNCH*	*SYNCH*	*CORRE*	*CORRE*
L_SA	-0.055***	-0.057***	-0.450**	-0.462**
	(-3.852)	(-3.946)	(-2.364)	(-2.415)
L_RP		-0.056**		-0.682**
		(-2.181)		(-2.107)
LNUM	0.141***	0.126***	1.153***	1.031***
	(7.128)	(6.301)	(4.448)	(3.895)
SIZE	0.373***	0.452***	5.332***	6.008***
	(23.917)	(22.241)	(27.081)	(23.178)
LMB		-0.146***		-1.198***
		(-6.846)		(-4.584)
LEV		0.171*		2.577**
		(1.731)		(2.113)
ROA		-0.042		-0.206
		(-0.592)		(-0.242)
NIND		-0.038		0.651
		(-0.367)		(0.515)
HERFSALE		0.413		-7.079
		(0.809)		(-1.074)
STDROA		0.017		0.595
		(0.148)		(0.455)
BIG4		0.062		0.619
		(1.502)		(1.220)
CONSTANT	-5.273***	-5.288***	3.369***	-0.644
	(-53.159)	(-9.408)	(2.706)	(-0.092)
Observations	39,568	39,568	39,568	39,568
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Adj. R^2	0.648	0.649	0.641	0.641

Table 3 presents the effect of Seeking Alpha coverage on stock return comovement. The sample contains 39,568 firm-year observations over the period 2004–2014. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm return and weekly market returns *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise; *T*-statistics robust to heteroscedasticity and clustered by firm are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Results relating to moderating effect of financial reporting opacity

$$Comovement_{i,t} = \alpha_0 + \beta_1 L\_SA_{i,t} + \beta_2 OPA_{i,t} + \beta_3 L\_SA_{i,t} * OPA_{i,t} + \beta_4 L\_RP_{i,t} + \beta_5 LNUM_{i,t} + \beta_6 SIZE_{i,t} + \beta_7 LMB_{i,t}$$
$$+ \beta_8 LEVERAGE_{i,t} + \beta_9 ROA_{i,t} + \beta_{10} NIND_{i,t} + \beta_{11} HERFSALE_{i,t} + \beta_{12} STDROA_{i,t} + \beta_{13} BIG4_{i,t} + \sum \alpha_i Firm_i + \sum \alpha_j Year_j + \varepsilon_{i,t}.$$

VARIABLES	*SYNCH*	*CORRE*	*SYNCH*	*CORRE*
L_SA	-0.048***	-0.201	-0.070***	-0.537*
	(-2.690)	(-0.842)	(-3.369)	(-1.835)
OPA	-0.014	-0.468		
	(-0.124)	(-0.397)		
OPA2			0.000***	0.000
			(3.641)	(1.028)
*L_SA*OPA*	-0.142***	-1.938***		
	(-4.297)	(-3.650)		
*L_SA*OPA2*			-0.073***	-0.451***
			(-4.711)	(-2.695)
L_RP	-0.050	-0.204	-0.231***	-3.413***
	(-1.451)	(-0.456)	(-4.297)	(-4.954)
LNUM	0.120***	1.008***	0.076**	0.395
	(4.531)	(3.006)	(2.315)	(0.867)
SIZE	0.419***	5.233***	0.391***	4.067***
	(15.145)	(14.989)	(10.664)	(8.301)
LMB	-0.141***	-0.818**	-0.153***	-0.863*
	(-4.988)	(-2.388)	(-4.144)	(-1.700)
LEV	0.172	3.186**	0.256*	3.324
	(1.347)	(1.990)	(1.669)	(1.635)
ROA	0.025	1.377	0.218	3.346**
	(0.285)	(1.304)	(1.567)	(1.963)
NIND	-0.328**	-0.004	-0.306*	1.542
	(-2.196)	(-0.002)	(-1.931)	(0.658)
HERFSALE	1.012	8.799	-0.344	-12.575
	(1.319)	(0.885)	(-0.434)	(-1.118)
STDROA	-0.006	2.047	-0.108	-1.749
	(-0.035)	(1.167)	(-0.444)	(-0.621)
BIG4	0.070	0.947	0.083	0.961
	(1.172)	(1.259)	(1.073)	(0.968)
CONSTANT	-3.621***	3.997	-2.135**	28.208**
	(-4.366)	(0.364)	(-2.350)	(2.165)
Observations	22,847	22,847	12,606	12,606
Firm fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Adj. R-squared	0.642	0.638	0.618	0.555

Table 4 presents the results of varying effect of Seeking Alpha coverage on stock return co-movement with respect to Opacity. The sample contains 22,847 (12,606) firm-year observations over the period 2004–2014 when *SYNCH* and *CORRE* are the dependent variable.[26] *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm return and weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *OPA* is the previous five years standard deviation of Accrual quality value based on Francis et al (2005) model; *OPA2* is the analyst earnings forecast dispersion scaled by the firm's opening stock price of the fiscal year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise. T-statistics robust to heteroscedasticity and clustered by firm are reported in the parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

[26] The sample size is reduced due to the availability of data required to calculate *OPA* and *OPA2*.

Table 5: Propensity score matching

VARIABLES	*SYNCH*	*CORRE*
L_SA	-0.053***	-1.085***
	(-3.013)	(-4.899)
L_RP	-0.055	-1.148**
	(-1.380)	(-2.347)
LNUM	0.085***	0.826**
	(3.045)	(2.363)
SIZE	0.377***	4.450***
	(13.243)	(12.650)
LMB	-0.119***	-0.929***
	(-4.222)	(-2.782)
LEVERAGE	0.177	3.165*
	(1.227)	(1.772)
ROA	0.004**	0.079***
	(2.141)	(3.667)
NIND	-0.046	-3.265*
	(-0.336)	(-1.691)
HERFSALE	-1.455**	-28.086***
	(-2.292)	(-3.675)
STDROA	-0.069***	-0.428***
	(-7.658)	(-3.955)
BIG4	0.012	-0.569
	(0.154)	(-0.609)
CONSTANT	-4.666***	33.606***
	(-5.725)	(3.087)
Observations	21,528	21,528
Firm fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Adj. R-squared	0.661	0.629

Table 5 presents the effect of Seeking Alpha coverage on stock return co-movement for PSM matched sample. The sample contains 21,528 firm-year observations over the period 2004–2014. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm return and weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals one if the firm is audited by one of the Big 4 audit firms, zero otherwise. Both firm and year fixed effects are used. T-statistics robust to heteroscedasticity and clustered by firm are reported in the parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 6: Two-stage-least-squares

VARIABLES	First stage	Second stage *SYNCH*	Second stage *CORRE*
L_SA_IND	0.029***		
	(11.683)		
L_ADX	0.037***		
	(2.919)		
L_SA_P		-0.479**	-11.620***
		(-2.470)	(-4.479)
L_RP	0.104***	-0.011	0.492
	(7.479)	(-0.361)	(1.143)
LNUM	0.034***	0.142***	1.440***
	(2.741)	(6.965)	(4.838)
SIZE	0.103***	0.498***	7.219***
	(10.453)	(17.552)	(18.942)
LMB	-0.029***	-0.160***	-1.586***
	(-2.767)	(-7.652)	(-5.678)
LEVERAGE	0.195***	0.255**	4.801***
	(4.478)	(2.552)	(3.587)
ROA	0.027	-0.032	0.047
	(1.037)	(-0.494)	(0.058)
NIND	0.168***	0.041	2.729**
	(3.290)	(0.400)	(2.015)
HERFSALE	0.768**	0.721	1.050
	(2.513)	(1.454)	(0.147)
STDROA	0.123**	0.072	2.046
	(2.569)	(0.662)	(1.568)
BIG4	0.053***	0.085**	1.219**
	(3.809)	(2.142)	(2.425)
Observations	38,821	38,821	38,821
Firm fixed effects	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Adj. R-squared	0.594	0.637	0.588

Table 6 presents the two-stage least squares (2SLS) results. We use two instrumental variables: L_ADX is the natural log of 1 plus the firm's annual advertisement expenditure; *L_SA_IND* is the natural log of 1 plus the total annual SA coverage of the industry to which the firm belongs. The sample contains 38,821 firm-year observations over the period 2004–2014. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry level return; *CORRE* is the percentage points of time series Pearson correlation coefficient between weekly firm return and weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_SA_P* is the predicted value of *L_SA* from the first stage. *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise. Both firm and year fixed effects are used. T-statistics robust to heteroscedasticity and clustered by firm are reported in the parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 7: "Day of the Week" effect

VARIABLES	*SYNCH*	*CORRE*
L_SA_WEEKENDS	-0.106***	-1.652***
	(-3.682)	(-5.192)
L_SA_WEEKDAY	-0.177***	-2.159***
	(-11.223)	(-13.469)
L_RP	-0.072***	-1.683***
	(-3.530)	(-9.457)
LNUM	0.116***	1.129***
	(4.727)	(5.061)
SIZE	0.796***	6.421***
	(32.523)	(31.468)
LMB	0.043**	0.109
	(2.162)	(0.656)
LEV	-0.212*	-0.815
	(-1.772)	(-0.801)
ROA	-0.001	0.102*
	(-0.082)	(1.820)
NIND	-0.168	-1.222
	(-1.568)	(-1.349)
HERFSALE	0.824	-6.953
	(1.416)	(-1.527)
STDROA	-0.026	-0.219
	(-1.193)	(-1.552)
BIG4	-0.075	0.375
	(-1.195)	(0.746)
Observations	108,333	112,346
firm fixed effects	Yes	Yes
time fixed effects	Yes	Yes
Adj. R-squared	0.665	0.675

Table 7 presents the effect of Seeking Alpha coverage on stock return co-movement using quarterly data. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-quarter estimation regressing daily stock return on daily market and industry level return; *CORRE* is the percentage points of time-series Pearson correlation coefficient between daily firm return and daily market returns; *L_SA_WEEKENDS* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles posted on Friday, Saturday and Sunday for a firm in the quarter; *L_SA_WEEKDAY* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles posted from Monday to Thursday for a firm in the quarter; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the quarter; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise. T-statistics robust to heteroscedasticity and clustered by firm are reported in the parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 8: Incorporation of future earnings into current stock price

VARIABLES	4(a) *Ret1*	4(b) *Ret2*	Wald Test of coefficient difference
EARNING	1.990***	5.046***	
	(9.445)	(16.364)	
L_SA	0.050***	0.087***	
	(7.876)	(8.565)	
*EARNING*L_SA*	0.006	0.051***	0.045***
	(1.626)	(6.284)	(23.64)
L_RP	0.003	-0.007***	
	(1.496)	(-2.608)	
*EARNING*L_RP*	0.025***	0.031***	
	(6.839)	(5.643)	
LNUM	0.054***	0.077***	
	(11.520)	(10.689)	
*EARNING*LNUM*	-0.050***	-0.095***	
	(-6.750)	(-6.724)	
SIZE	-0.077***	-0.150***	
	(-25.294)	(-33.574)	
*EARNING*SIZE*	0.061***	0.088***	
	(7.543)	(7.258)	
LMB	-0.162***	-0.225***	
	(-29.436)	(-27.377)	
*EARNING*LMB*	0.016***	0.035***	
	(3.279)	(4.547)	
LEVERAGE	0.340***	0.573***	
	(12.572)	(14.114)	
*EARNING*LEVERAGE*	0.932***	1.915***	
	(6.989)	(9.663)	
ROA	-0.061***	-0.165***	
	(-5.502)	(-7.588)	
*EARNING*ROA*	-0.000	0.001**	
	(-1.425)	(1.972)	
NIND	0.000	0.008	
	(0.000)	(0.986)	
EARNING NIND*	-0.368***	-0.855***	
	(-11.774)	(-18.392)	
HERFSALE	-0.074	0.059	
	(-0.922)	(0.482)	
*EARNING*HERFSALE*	0.182	-3.462***	
	(0.304)	(-3.835)	
STDROA	0.053***	0.055***	
	(4.967)	(4.301)	
*EARNING*STDROA*	-0.000	-0.004***	
	(-0.552)	(-4.362)	
BIG4	0.134***	0.319***	
	(12.043)	(19.399)	
*EARNING*BIG4*	0.039*	0.133***	
	(1.656)	(3.793)	
CONSTANT	1.492***	1.978***	
	(38.805)	(33.987)	
Observations	34,866	34,866	
Adj. R-squared	0.027	0.087	

Table 8 presents the results on whether firms with higher Seeking Alpha coverage incorporate future earnings into their prices more efficiently. We estimate the system of Equations 10(a) and 10(b) using Seemingly Unrelated Regression (SUR). *RET1* is the return from the previous year and *RET2* is the previous two years' return; EARNINGS is the income before extraordinary items divided by total assets; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals 1 if the firm is audited by one of the Big 4 audit firms, zero otherwise. We report the Wald test results on the difference of coefficients on EARNINGS*L_SA between two equations. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table 9: Baseline results across size groups

SIZE GROUP VARIABLES	Large *SYNCH*	Small *SYNCH*	Large *CORRE*	Small *CORRE*
L_SA	-0.046***	-0.052*	-0.500**	-1.034***
	(-3.145)	(-1.765)	(-2.402)	(-3.053)
L_RP	-0.165***	0.032	-2.580***	0.787*
	(-5.855)	(0.854)	(-6.912)	(1.813)
LNUM	0.058***	0.247***	0.346	2.635***
	(2.898)	(6.645)	(1.265)	(5.749)
SIZE	0.268***	0.537***	1.935***	7.891***
	(9.606)	(18.959)	(5.342)	(23.216)
LMB	-0.055**	-0.193***	0.153	-2.292***
	(-2.145)	(-6.608)	(0.439)	(-6.849)
LEV	-0.046	0.212	0.826	1.694
	(-0.412)	(1.464)	(0.543)	(1.030)
ROA	0.189	-0.156**	0.939	-2.094**
	(1.511)	(-2.023)	(0.583)	(-2.310)
NIND	-0.674***	0.522***	-3.503**	4.547**
	(-6.639)	(3.421)	(-2.206)	(2.552)
HERFSALE	0.393	0.487	-12.996*	0.940
	(0.718)	(0.555)	(-1.675)	(0.088)
STDROA	0.061	-0.023	0.274	1.207
	(0.344)	(-0.163)	(0.124)	(0.805)
BIG4	0.059	0.058	-0.337	-0.003
	(0.970)	(1.168)	(-0.432)	(-0.005)
Observations	19,220	19,036	19,220	19,036
Firm clustering	Yes	Yes	Yes	Yes
Time clustering	Yes	Yes	Yes	Yes
Adj. R-squared	0.555	0.421	0.488	0.500

Table 9 presents results comparing the effect of Seeking Alpha coverage on co-movement among small and large firms. The entire sample is partitioned into high and low size sub-sample based on the annual median values of *SIZE*. *SYNCH*, stock price synchronicity, is defined as the log-transformation of the adjusted R^2 of the firm-year estimation regressing weekly stock return on weekly market- and industry-level returns; *CORRE* is the percentage points of time-series Pearson correlation coefficient between weekly firm returns and the weekly market returns; *L_SA* is the natural log of 1 plus the number of single-ticker Seeking Alpha articles for a firm in the year; *L_RP* is the natural log of 1 plus the number of most relevant firm-specific news articles from RavenPack Analytics for a firm in the year; *LNUM* is the natural log of 1 plus the number of analysts following the firm for the previous fiscal year; *SIZE* is the natural log of the firm's market capitalization at the end of the last fiscal year; *LMB* is the natural log of market capitalization scaled by the book value of equity at the end of the last fiscal year; *LEVERAGE* is the total long-term debt scaled by total assets at the end of the last fiscal year; *ROA* is income before extraordinary items divided by total assets at the end of the last fiscal year; *NIND* is the natural log of the number of firms in the industry to which firm i belongs at the end of the last fiscal year; *HERFSALE* is the sum of the squared terms of the proportion of a firm's revenue to total revenue in the industry at the end of the last fiscal year; *STDROA* is the standard deviation of the ratio between income before extraordinary items and total assets in the previous five years; *BIG4* is a dummy variable that equals one if the firm is audited by one of the Big 4 audit firms, zero otherwise. T-statistics robust to heteroscedasticity and clustered by firm and year are reported in the parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.